

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-065590

(43)Date of publication of application : 09.03.1999

(51)Int.Cl.

G10L 3/00
G10L 3/00
H04Q 7/38
H04M 1/274
H04M 3/42

(21)Application number : 09-228567

(71)Applicant : NEC CORP

(22)Date of filing : 25.08.1997

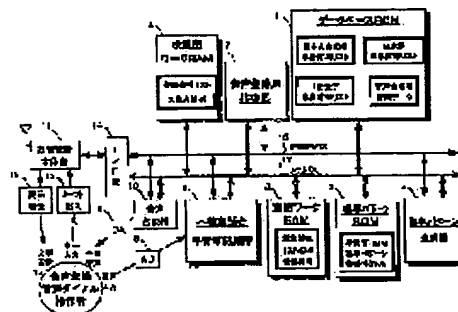
(72)Inventor : Tabei Kazuhiro

(54) VOICE RECOGNITION DIALING DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To reduce cumbersome key operations for registering names and corresponding telephone numbers on a portable telephone by outputting the recognition result of the names and the telephone numbers by the recognition processing, which uses the monosyllable code data stream of the registered names or the telephone numbers with respect to input voices and standard pattern data in half syllable units.

SOLUTION: A voice registration means, such as a general speaker half syllable voice recognizer 6 registers the uttered voices of names and telephone numbers through the voice input of the object names and the corresponding telephone numbers for a voice recognition dialing. The dialing voice means of the recognizer 6 conducts a dialing, using the monosyllabic code data column of the names and the telephone number which have been registered beforehand from the voice analog signals of the names and the telephone numbers. The recognizer 6 conducts the recognition process, which uses the monosyllabic code data column of the names and the telephone numbers that have been registered with respect to input voice and the standard pattern data in half syllable units. Then, the name of the telephone number monosyllabic code data column of the candidate having a small cumulative distance value is outputted as the recognized result.



LEGAL STATUS

[Date of request for examination] 25.08.1997

[Date of sending the examiner's decision of rejection] 10.05.2000

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3447521

[Date of registration]	04.07.2003
[Number of appeal against examiner's decision of rejection]	2000-08467
[Date of requesting appeal against examiner's decision of rejection]	08.06.2000
[Date of extinction of right]	

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-65590

(43)公開日 平成11年(1999) 3月9日

(51)Int.Cl. ⁶	識別記号	F I
G 1 0 L 3/00	5 5 1	G 1 0 L 3/00 5 5 1 A
	5 2 1	5 2 1
H 0 4 Q 7/38		H 0 4 M 1/274
H 0 4 M 1/274		3/42 P
3/42		H 0 4 B 7/26 1 0 9 Q
審査請求 有 請求項の数9 OL (全 25 頁)		

(21)出願番号 特願平9-228567

(22)出願日 平成9年(1997) 8月25日

(71)出願人 000004237

日本電気株式会社
東京都港区芝五丁目7番1号

(72)発明者 田部井 和彦

東京都港区芝五丁目7番1号 日本電気株
式会社内

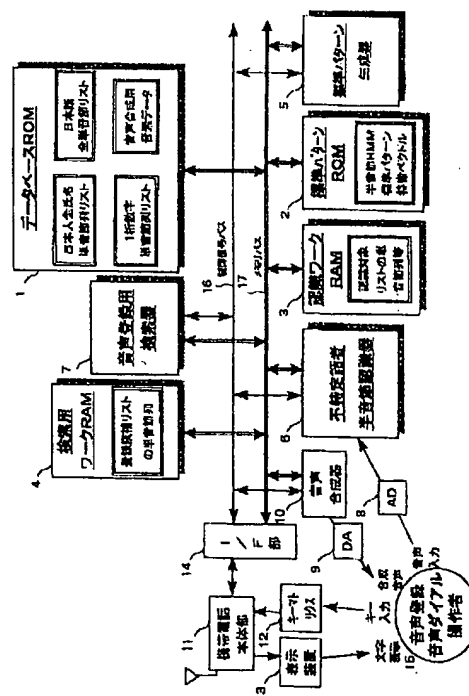
(74)代理人 弁理士 若林 忠 (外4名)

(54)【発明の名称】 音声認識ダイヤル装置

(57)【要約】

【課題】 携帯電話機等における氏名と電話番号登録操作、およびダイヤル操作等のキー操作における煩雑さを低減する音声認識ダイヤル装置を提供する。

【解決手段】 日本人全氏名のデータベースを単音節コード単位（ひらがな文字コード；ASCIIあるいはSJIS等）であらかじめデータベースとしてROM等に格納しておき、これと単音節単位で認識処理可能な不特定話者半音節音声認識器により、氏名および電話番号を単音節単位で離散発声した音声信号から自動登録させることによる音声登録機能と、登録された氏名および電話番号の単音節コードを認識対象とした不特定話者半音節音声認識器による認識処理により、音声ダイヤルを行う機能とにより、キー操作を低減できる。



【特許請求の範囲】

【請求項1】 氏名と電話番号を発声した音声のPCM信号から音声認識処理により単音節コードデータ列を取得し、音声認識ダイアル用の対象氏名と電話番号を音声入力で登録する音声登録手段と、

氏名又は電話番号の音声アナログ信号からあらかじめ登録済みの氏名と電話番号の単音節コードデータ列を用いてダイアルをする音声ダイアル手段と、

入力音声に対する、登録済み氏名又は電話番号の単音節コードデータ列と半音節単位の標準パターンデータとを用いた認識処理により、累積距離値の小さい候補の氏名又は電話番号の単音節コードデータ列を認識結果として出力する不特定話者半音節音声認識器とを備えている音声認識ダイアル装置。

【請求項2】 前記音声登録手段は、音声アナログ信号をPCM信号へ変換するADコンバータと、前記PCM信号を前記音声アナログ信号へ変換するDAコンバータと、

日本語の1音節を表す単音節コードデータを受信しひらがなと数字と漢字の表記文字を表示する表示手段と、

前記単音節コードデータを受信し音声PCMデータに変換しDAコンバータに出力する音声応答手段と、

日本人氏名と日本語の全単音節と数字1桁の各々について発音表記の単音節列を単音節コードデータで格納したデータベースROMと、

該データベースROM内の氏名項目又は単音節項目又は1桁数字項目の何れかの項目に属する1氏名又は1単音節又は1桁数字の何れかの1リストに対する単音節コードデータ列を読み出し認識ワークRAMへ格納する手段と、

単音節単位又は数字1桁単位に発声された前記音声アナログ信号をAD変換した前記PCM信号を前記不特定話者半音節音声認識器へ入力し、前記単音節単位又は数字1桁単位の認識処理を行い第1候補あるいは複数候補の認識結果を得る手段と、

前記単音節単位又は数字1桁単位の音声認識結果情報と前記データベースROM内の氏名項目又は1桁数字項目の単音節コードデータ列との両方の情報を用いて発声内容に最も近いと推測される氏名あるいは数字1桁以上の単音節コードデータ列を検索し出力する音声登録用検索器と、

検索された氏名又は電話番号の単音節コードデータ列を各氏名と電話番号を対応付けて複数の該各氏名と電話番号リストを認識ワークRAMへ蓄積格納する手段とを備えている請求項1に記載の音声認識ダイアル装置。

【請求項3】 前記音声登録用検索器は、前記不特定話者半音節認識器から前記単音節単位又は数字1桁単位の音声認識結果情報を累積距離値と共に情報受信し、前記データベースROMに格納されている氏名

項目又は1単音節又は1桁数字項目の何れかの項目に属する1氏名又は1単音節又は1桁数字の何れかの1リスト内で、1単音節単位に発声された音声信号から最も距離の近い単音節コードデータ列を検索し認識ワークRAMへ格納する手段を有する請求項2に記載の音声認識ダイアル装置。

【請求項4】 前記音声ダイアル手段は、ダイアルの宛先の氏名又は電話番号に対応する発声した音声アナログ信号をADコンバータで変換したPCM信号を前記不特定話者半音節音声認識器へ入力し、認識ワークRAM内の登録済み氏名又は電話番号リストに対して前記発声した音声アナログ信号に最も近いと推測される認識結果を単音節コードデータ列で取得する手段と、認識結果が氏名の場合は対応付けされた電話番号の単音節コードデータ列を前記認識ワークRAM内から検索して出力する手段と、

前記電話番号の単音節コードデータ列から電話端末本体への電話発呼信号へ変換する手段とを有する請求項1に記載の音声認識ダイアル装置。

【請求項5】 前記不特定話者半音節音声認識器は、認識ワークRAM又はデータベースROMに格納されている氏名又は電話番号の単音節単位の単音節コードデータ列に対し半音節単位の半音節コードデータ列へ変換する手段と、

氏名又は電話番号の前記半音節コードデータ列に対して標準パターン生成器の不特定話者半音節音声認識装置用の標準パターンデータ生成により得られた標準パターンを格納した標準パターンROM内からどの半音節単位の隠れマルコフモデルが含まれているかを調べ、さらに前記隠れマルコフモデル状態の連結を示す半音節隠れマルコフモデル状態コードデータ列へ変換する手段と、

氏名又は電話番号の前記半音節隠れマルコフモデル状態コードデータ列を氏名-電話番号の関係で対応付けし、さらに他の氏名-電話番号リストと識別可能なように番号付きリストに変換して前記認識ワークRAMへ格納する手段と、

音声アナログ信号をAD変換した音声PCM信号からフレーム単位の入力特徴ベクトルを抽出する音声分析特徴抽出器と、

抽出された前記入力特徴ベクトルを前記認識ワークRAMへ格納する手段と、

前記入力特徴ベクトルと、標準半音節隠れマルコフモデルパターンROMに格納されている全半音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を算出する状態距離計算器と、

算出された前記状態距離値に番号付けて認識ワークRAMに格納する手段と、

前記認識ワークRAMに格納された認識対象の各氏名又は電話番号に対する前記半音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と前記標

準パターンROM内に格納されている状態間遷移距離値とを用いてフレーム同期Viterbiアルゴリズムにより入力音声時間長分の全フレームに対する累積距離値を算出する累積状態距離計算器と、

最も前記累積距離値の小さい第1候補又は第1～第N候補の氏名又は電話番号に対する単音節コードデータ列を認識結果として出力する手段とを有する請求項1から請求項4の何れか1項に記載の音声認識ダイアル装置。

【請求項6】 前記標準パターン生成器は、統計的に必要とされる人数分の多数話者の音声アナログ信号をAD変換したPCM信号を、波形表示あるいは試験等により所定数の種類の単音節単位のPCM信号へ分割する手段と、

前記単音節単位に分割された全てのPCM信号をバッファリングし、波形表示あるいは試験等により所定数の種類の単音節単位のPCMデータ信号へ分割する手段と、前記単音節単位に分割された全ての各PCM信号をフレーム単位に分割する手段と、

前記フレーム単位に分割された全てのPCM信号に対して特徴ベクトルを抽出する音声分析特徴抽出器と、状態数が所定の個数の隠れマルコフモデルにおいて前記所定の個数の状態出力確率関数と前記所定の個数の2倍個数分の状態遷移確率のパラメータを初期値設定する手段と、

前記状態出力確率関数を初期設定する際に多次元正規分布確率密度関数を用いると共に、母数として平均ベクトルおよび共分散行列の各成分を初期値設定する手段と、所定の種類分の単音節毎に得られた前記統計的に必要とされる人数分のフレーム単位の前記特徴ベクトルから、各単音節毎に前記統計的に必要とされる人数分の特徴ベクトルサンプルとして整理して、Forward-Backwardアルゴリズムという反復的手法により前記所定の個数の平均ベクトルおよび共分散行列の各成分値と前記所定の個数の2倍個数分の状態遷移確率とを得る手段と、

所定の種類分の単音節毎に得られた隠れマルコフモデルのパラメータ群である多次元正規分布確率密度関数の平均ベクトルと共分散行列と状態遷移確率とのパラメータ値を標準パターンデータとして標準パターンROMへ格納する手段とを有する請求項5に記載の音声認識ダイアル装置。

【請求項7】 前記音声分析特徴抽出器は、入力音声のPCM信号をフレーム分割したフレームPCM信号をプリエンファシス処理することにより高周波数帯域を強調する手段と、前記プリエンファシス処理済みフレームPCM信号に対し窓処理することによりこの後のFFT処理のためのフレーム境界のスムージング処理をする手段と、

前記窓処理後のフレームPCM信号をN次FFT変換処理することにより線形周波数軸上のN次複素係数ベクトルへ変換する手段と、

該N次複素係数ベクトルから複素数の絶対値計算によりN次振幅係数ベクトルへ変換する手段と、

該N次振幅係数ベクトルに対し対数演算を施してN次対数振幅係数ベクトルを算出する手段と、

該N次対数振幅係数ベクトルに対し時間軸上への逆離散余弦変換によりP次ケプストラムベクトルを算出する手段と、

該P次ケプストラムベクトルの高時間成分を抑圧する処理により、声道特性とピッチ特性を分離し声道特性のみを抽出したP次声道特性ケプストラムベクトルを算出する手段と、

該P次声道特性ケプストラムベクトルに対し周波数軸上への離散余弦変換を行い、線形周波数軸上におけるN次声道特性対数振幅係数ベクトルに変換する手段と、

該N次声道特性対数振幅係数ベクトルに対しメル周波数軸上における等分割点上のスペクトル成分へ補間あるいはスムージング処理を施したベクトル成分を算出することにより人間の聴覚周波数分解能特性に合わせたN次声道特性対数振幅メル尺度係数ベクトルへ変換する手段と、

該N次声道特性対数振幅メル尺度係数ベクトルに対し時間軸上への逆離散余弦変換によりQ次メルケプストラムベクトルへ変換する手段とを有する請求項5又は請求項6に記載の音声認識ダイアル装置。

【請求項8】 前記状態距離計算器は、入力音声のアナログ信号をAD変換しさらにフレーム分割したフレームPCM信号から前記音声分析特徴抽出器により得られた入力特徴ベクトルと標準単音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を所定の距離計算式により算出する手段とを有する請求項5に記載の音声認識ダイアル装置。

【請求項9】 前記累積距離計算器は、認識対象の氏名又は電話番号に対する単音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と標準パターンROM内に格納されている状態間遷移距離値とを用いてフレーム同期Viterbiアルゴリズムにより入力音声の全フレームに対する累積距離値を算出する手段とを有する請求項5に記載の音声認識ダイアル装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、音声認識ダイアル装置に関する。

【0002】

【従来の技術】 従来では、認識単語登録の作業には使用者のキー入力装置への操作が必要であった。例えば、氏名と電話番号の1人分の登録を実行しようとするには、使用者が[人名の平仮名文字数] + [電話番号桁数] 分のキー操作入力が必要となる。

【0003】携帯電話では小型化のために、パソコンのような全文字分の入力キーを持った入力装置ではなく、入力キーの数量が限られているため、複数文字分が重複されて割り当てられている。

【0004】例えば、携帯電話の入力装置を例に、人名＝7文字、電話番号＝10桁を登録しようとする場合を考える。

【0005】携帯電話では、ア行＝1キー、力行＝1キー～ラ行＝1キー、ワラン＝1キーという形でキーが割り当てられており、7文字を登録する場合最短で1×7＝7回のキー入力、最長で5×7＝35回のキー入力、平均で2.5×7＝17.5回のキー入力となる。

【0006】数字は全部がキーに割り当てられており、操作モード変更することで入力できるため、1+10回のキー入力となる。

【0007】以上の例では、平均操作時間を1キー入力当たり＝2秒とすると2秒×(17.5+10)回＝55秒となる。

【0008】

【発明が解決しようとする課題】上述した従来の技術では、音声認識ダイアル用の氏名と電話番号の登録操作時間が長く、また操作間違いも少なくない。

【0009】また音声を用いた従来技術に特開平2-135847号公報に開示された音声応答認識自動ダイアル電話機があるが、この従来技術では、音声入力信号を認識し文字データに、変換する機能を用いているが、実現手段が明記されていない。

【0010】本発明の目的は、携帯電話機等における氏名と電話番号登録操作、およびダイアル操作等のキー操作における煩雑さを低減する音声認識ダイアル装置を提供することにある。

【0011】

【課題を解決するための手段】本発明の音声認識ダイアル装置は、氏名と電話番号を発声した音声のPCM信号から音声認識処理により単音節コードデータ列を取得し、音声認識ダイアル用の対象氏名と電話番号を音声入力に登録する音声登録手段と、氏名又は電話番号の音声アナログ信号からあらかじめ登録済みの氏名と電話番号の単音節コードデータ列を用いてダイアルをする音声ダイアル手段と、入力音声に対する、登録済み氏名又は電話番号の単音節コードデータ列と半音節単位の標準パターンデータとを用いた認識処理により、累積距離値の小さい候補の氏名又は電話番号の単音節コードデータ列を認識結果として出力する不特定話者半音節音声認識器とを備えている。

【0012】また、音声登録手段は、音声アナログ信号をPCM信号へ変換するADコンバータと、PCM信号を音声アナログ信号へ変換するDAコンバータと、日本語の1音節を表す単音節コードデータを受信しひらがなと数字と漢字の表記文字を表示する表示手段と、単音節

コードデータを受信し音声PCMデータに変換しDAコンバータに出力する音声応答手段と、日本人氏名と日本語の全単音節と数字1桁の各々について発音表記の単音節列を単音節コードデータで格納したデータベースROMと、データベースROM内の氏名項目又は単音節項目又は1桁数字項目の何れかの項目に属する1氏名又は1単音節又は1桁数字の何れかの1リストに対する単音節コードデータ列を読出し認識ワークRAMへ格納する手段と、単音節単位又は数字1桁単位に発声された音声アナログ信号をAD変換したPCM信号を不特定話者半音節音声認識器へ入力し、単音節単位又は数字1桁単位の認識処理を行い第1候補あるいは複数候補の認識結果を得る手段と、単音節単位又は数字1桁単位の音声認識結果情報とデータベースROM内の氏名項目又は1桁数字項目の単音節コードデータ列との両方の情報を用いて発声内容に最も近いと推測される氏名あるいは数字1桁以上の単音節コードデータ列を検索し出力する音声登録用検索器と、検索された氏名又は電話番号の単音節コードデータ列を各氏名と電話番号を対応付けて複数の各氏名と電話番号リストを認識ワークRAMへ蓄積格納する手段とを備えていてもよい。

【0013】また、音声登録用検索器は、不特定話者半音節認識器から単音節単位又は数字1桁単位の音声認識結果情報を累積距離値と共に情報受信し、データベースROMに格納されている氏名項目又は1単音節又は1桁数字項目の何れかの項目に属する1氏名又は1単音節又は1桁数字の何れかの1リスト内で、1単音節単位に発声された音声信号から最も距離の近い単音節コードデータ列を検索し認識ワークRAMへ格納する手段を有してもよい。

【0014】また、音声ダイアル手段は、ダイアルの宛先の氏名又は電話番号に対応する発声した音声アナログ信号をADコンバータで変換したPCM信号を不特定話者半音節音声認識器へ入力し、認識ワークRAM内の登録済み氏名又は電話番号リストに対して発声した音声アナログ信号に最も近いと推測される認識結果を単音節コードデータ列で取得する手段と、認識結果が氏名の場合は対応付けされた電話番号の単音節コードデータ列を認識ワークRAM内から検索して出力する手段と、電話番号の単音節コードデータ列から電話端末本体への電話発呼信号へ変換する手段とを有してもよい。

【0015】また、不特定話者半音節音声認識器は、認識ワークRAM又はデータベースROMに格納されている氏名又は電話番号の単音節単位の単音節コードデータ列に対し半音節単位の半音節コードデータ列へ変換する手段と、氏名又は電話番号の半音節コードデータ列に対して標準パターン生成器の不特定話者半音節音声認識装置用の標準パターンデータ生成により得られた標準パターンを格納した標準パターンROM内からどの半音節単位の隠れマルコフモデルが含まれているかを調べ、さら

に隠れマルコフモデル状態の連結を示す半音節隠れマルコフモデル状態コードデータ列へ変換する手段と、氏名又は電話番号の半音節隠れマルコフモデル状態コードデータ列を氏名-電話番号の関係で対応付けし、さらに他の氏名-電話番号リストと識別可能なように番号付きリストに変換して認識ワークRAMへ格納する手段と、音声アナログ信号をAD変換した音声PCM信号からフレーム単位の入力特徴ベクトルを抽出する音声分析特徴抽出器と、抽出された入力特徴ベクトルを認識ワークRAMへ格納する手段と、入力特徴ベクトルと、標準半音節隠れマルコフモデルパターンROMに格納されている全半音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を算出する状態距離計算器と、算出された状態距離値に番号付けして認識ワークRAMに格納する手段と、認識ワークRAMに格納された認識対象の各氏名又は電話番号に対する半音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と標準パターンROM内に格納されている状態間遷移距離値とを用いてフレーム同期Viterbialgorithmにより入力音声時間長分の全フレームに対する累積距離値を算出する累積状態距離計算器と、最も累積距離値の小さい第1候補又は第1～第N候補の氏名又は電話番号に対する半音節コードデータ列を認識結果として出力する手段とを有してもよい。

【0016】また、標準パターン生成器は、統計的に必要とされる人数分の多数話者の音声アナログ信号をAD変換したPCM信号を、波形表示あるいは試聴等により所定数の種類の半音節単位のPCM信号へ分割する手段と、半音節単位に分割された全てのPCM信号をバッファリングし、波形表示あるいは試聴等により所定数の種類の半音節単位のPCMデータ信号へ分割する手段と、半音節単位に分割された全ての各PCM信号をフレーム単位に分割する手段と、フレーム単位に分割された全てのPCM信号に対して特徴ベクトルを抽出する音声分析特徴抽出器と、状態数が所定の個数の隠れマルコフモデルにおいて所定の個数の状態出力確率関数と所定の個数の2倍個数分の状態遷移確率のパラメータを初期値設定する手段と、状態出力確率関数を初期設定する際に多次元正規分布確率密度関数を用いると共に、母数として平均ベクトルおよび共分散行列の各成分を初期値設定する手段と、所定の種類分の半音節毎に得られた統計的に必要とされる人数分のフレーム単位の特徴ベクトルから、各半音節毎に統計的に必要とされる人数分の特徴ベクトルサンプルとして整理して、Forward-Backward algorithmという反復的手法により所定の個数の平均ベクトルおよび共分散行列の各成分値と所定の個数の2倍個数分の状態遷移確率とを得る手段と、所定の種類分の半音節毎に得られた隠れマルコフモデルのパラメータ群である多次元正規分布確率密度関数の平均ベクトルと共分散行列と状態遷移確率とのパラメータ値を

標準パターンデータとして標準パターンROMへ格納する手段とを有してもよい。

【0017】さらに、音声分析特徴抽出器は、入力音声のPCM信号をフレーム分割したフレームPCM信号をプリエンファシス処理することにより高周波数帯域を強調する手段と、プリエンファシス処理済みフレームPCM信号に対し窓処理することによりこの後のFFT処理のためのフレーム境界のスムージング処理をする手段と、窓処理後のフレームPCM信号をN次FFT変換処理することにより線形周波数軸上のN次複素係数ベクトルへ変換する手段と、N次複素係数ベクトルから複素数の絶対値計算によりN次振幅係数ベクトルへ変換する手段と、N次振幅係数ベクトルに対し対数演算を施してN次対数振幅係数ベクトルを算出する手段と、N次対数振幅係数ベクトルに対し時間軸上への逆離散余弦変換によりP次ケプストラムベクトルを算出する手段と、P次ケプストラムベクトルの高時間成分を抑圧する処理により、声道特性とピッチ特性を分離し声道特性のみを抽出したP次声道特性ケプストラムベクトルを算出する手段と、P次声道特性ケプストラムベクトルに対し周波数軸上への離散余弦変換を行い、線形周波数軸上におけるN次声道特性対数振幅係数ベクトルに変換する手段と、N次声道特性対数振幅係数ベクトルに対しメル周波数軸上における等分割点上のスペクトル成分へ補間あるいはスムージング処理を施したベクトル成分を算出することにより人間の聴覚周波数分解能特性に合わせたN次声道特性対数振幅メル尺度係数ベクトルへ変換する手段と、N次声道特性対数振幅メル尺度係数ベクトルに対し時間軸上への逆離散余弦変換によりQ次メルケプストラムベクトルへ変換する手段とを有してもよい。

【0018】さらに、状態距離計算器は、入力音声のアナログ信号をAD変換しさらにフレーム分割したフレームPCM信号から音声分析特徴抽出器により得られた入力特徴ベクトルと標準半音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を所定の距離計算式により算出する手段とを有してもよい。

【0019】さらに、累積距離計算器は、認識対象の氏名又は電話番号に対する半音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と標準パターンROM内に格納されている状態間遷移距離値とを用いてフレーム同期Viterbialgorithmにより入力音声の全フレームに対する累積距離値を算出する手段とを有してもよい。

【0020】従って、本発明により、音声認識ダイアル用の氏名と電話番号の登録操作時間が従来例と比べて短縮され、また操作間違いも少なくなる。

【0021】また、従来技術の認識自動ダイアル電話機では、明記されていなかった音声から文字表示を行う音声認識処理について、この技術の実現手段を明記することで本発明の属する技術分野において実際に利用可能と

なる。

【0022】

【発明の実施の形態】本発明の実施の形態について図面を参照して説明する。図1は本発明の実施の形態の全体構成を示すブロック図である。まず、記憶装置として2種類のROM（データベースROM1、標準パターンROM2）と、2種類のRAM（認識ワークRAM3、検索ワークRAM4）とがある。

【0023】また、機能ブロックとして標準パターン生成器5と、不特定話者半音節音声認識器6と、音声登録用検索器7と、ADコンバータ8と、DAコンバータ9と、音声合成器10と、携帯電話本体部11と、キーマトリクス12と、表示装置13と、I/F部14とがある。以降では、音声登録動作と音声ダイヤル動作に分けて実施内容を説明する。また、これに続き主要な機能ブロックの内部詳細動作を、標準パターン生成器5と、不特定話者半音節音声認識器6と、音声登録用検索器7の、各々について各記憶装置間との連係動作内容も含めて説明する。

【0024】また、図1に示す機能ブロックに対する実施の形態としては、基本的に半導体集積回路と複合装置等により実現可能である。まずROM1、2と、RAM3、4と、ADコンバータ8と、DAコンバータ9などは、半導体集積回路となる。また、携帯電話本体11と、キーマトリクス12と、表示装置13（これはLCD等）と、I/F部14（これは携帯電話と拡張機器を接続するための拡張コネクタが利用可）は、複合装置となる。さらに、音声登録用検索器7と、不特定話者半音節音声認識器6と、音声合成器10は、CPUあるいは音声信号処理を高速演算可能なDSPと呼ばれるマイクロプログラム内蔵可能な半導体集積回路上におけるソフトウェアにより実現することが可能である。

【0025】1. 音声登録動作

まず、データベースROM1に格納されている全ての単音節単位（ひらがな1文字、即ち50音+α（濁音、拗音等））の単音節コードデータ（ASCIIあるいはJISあるいはSJIS等）を読み出し、更に各単音節コードデータを半音節コードデータ（半音節単位とは1音節をさらに半分に分割した音素単位をいい、例“た”の場合、TA→Tー、ーAに分割する）に変換して認識ワークRAM3へ格納する。なお、単音節→半音節への変換は後述の3. 不特定話者半音節音声認識器に内蔵された機能である。

【0026】次に登録する氏名と電話番号を離散単音節単位あるいは数字1桁単位に離散発声することにより、この登録氏名と登録電話番号のひらがな文字情報として単音節コードデータを、また数字情報として数字コードデータを以下に詳細に述べる（1）氏名登録動作および（2）電話番号登録動作により得ることができる。さらに、得られた氏名と電話番号の対応関連情報と、登録順

の番号を付けて認識ワークRAM3へ格納しておく。これにより、一連の音声登録動作が終了する。以降の（1）および（2）では、氏名を“たかはし”、電話番号を“03-123-4567”という例を用いて、音声登録動作の詳細内容について説明する。

【0027】（1）氏名登録動作

図2は音声登録動作モードのフローチャートである。

【0028】まず発声者は、“た”：“か”：“は”：“し”と単音節毎に無音間隔を入れて発声した音声アナログ信号をADコンバータ8によりPCM信号へ変換する（S3）。これらの4つの単音節分の各PCM信号に対して不特定話者半音節音声認識器6により、単音節単位の認識結果を得る（S4）。ここで認識結果は、近いものから第1～第5候補まで出力されるものとする。なお、複数の単音節認識結果候補を必要とする理由は、半音節等の音素単位の音声認識器では通常最小音素単位の認識精度が低く、例えば“た”と発声しても第1候補には同母音系の音節として、“あかさたなはまやらわ”

（拗音、濁音等の同母音系も含む）のような認識結果が出力される確率が高いため、第1候補だけではほとんど正確な認識結果が得られないのである。そこで認識結果として複数候補を用いるならば真の発声音節の認識結果がこれらの複数候補に含まれる可能性が高くなり、さらに、次に述べる登録氏名決定のための音声登録用検索器7とデータベースROM1の情報とから認識精度が上がることになる。

【0029】さて、氏名例の“た”“か”“は”“し”という4音節の離散発声アナログ信号をAD変換した各々のPCM信号に対して、全ての単音節を認識候補とした（実際には、データベースROM1から認識ワークRAM3へ全ての単音節コードデータを転送しておく）不特定話者半音節音声認識器6により図3に示すように4音節×5候補分の認識結果を得る。

【0030】これらの情報は4. 音声登録用検索器で説明するように登録ワークRAMへ格納される。

【0031】次に音声登録用検索器7により、これらの認識結果とデータベースROM1内の4音節に限定した（発声回数が4回のため）氏名リストから検索処理により登録氏名を決定する（S9）。

【0032】なお、氏名の検索結果を複数候補とする場合には、表示機能あるいは音声合成機能等を用いて、キー入力等により最終的に利用者（＝発声者）選択させることも出来る（S10）。

【0033】（2）電話番号登録動作

図4は電話番号登録動作モードのフローチャートである。

【0034】電話番号登録は、認識ワークRAM3に数字の単音節結合リスト「〃 ぜ+ろ（0）〃 ～〃 き+ゆ+う（9）〃」をあらかじめロードしておき、電話番号の桁数分だけ数字1桁ずつ離散発声を行うことにより（S

13) 得られる認識結果を図5に示すように数字を直接表すコードに変換する(S14)。なお、数字認識の場合は、単音節認識に比べ認識対象リストも10程度であり、また2音節以上の認識の場合は、認識精度も上がるため、あらかじめ電話番号リストをデータベースROM1に準備する必要がない。数字の結果を複数候補とする場合には、表示機能あるいは音声合成機能等を用いて、キー入力等により最終的に利用者(=発声者)に選択させることも出来る(S16)。

【0035】なお、電話番号の登録の際は、携帯電話等においてはキー入力操作も選択可能にしておいてもよい。一般に数字のキー割り当ては、1桁ずつある場合が多いので氏名入力ほど煩雑さの程度が低いからである。

【0036】2. 音声ダイアル動作

図6は音声ダイアル動作モードのフローチャートである。

【0037】ダイアルを始める前に氏名または電話番号の音声入力選択をキー入力等の指定により、認識ワークRAM3内に格納されている氏名あるいは電話番号のいずれを認識対象とするかの初期設定を行っておく(S22、S23、S30)。

【0038】次に氏名あるいは電話番号の発声を行い(S24、S31)、このアナログ音声信号をAD変換したPCM信号に対して、不特定話者半音節音声認識器により認識処理を行う(S25、S32)。

【0039】この認識結果は、初期設定において、氏名なのか、それとも電話番号であるかがわかっている。そのため電話番号の認識結果を得た場合は、その情報により電話番号の数字コード等からダイアル用の発呼信号へ変換してダイアルを実行できる(S36、S37)。氏名の認識結果を得た場合には、氏名と電話番号の関係付けされた情報から電話番号の数字コードを特定して同様にダイアルが可能となる(S29、S36、S37)。なお、電話番号あるいは氏名の認識処理後にダイアルを実行する前に氏名の文字と電話番号の数字を表示したり、あるいは音声合成器により音を出力したりすることにより1回の確認手続きを入れたりする(S27、S34)ことで、より親和性のある音声ダイアル機能にすることも可能である。

【0040】3. 不特定話者半音節音声認識動作

(1) 標準パターン生成処理

図7及び図8は標準パターン生成器の機能ブロック構成図である。

【0041】図9は標準パターン生成器による標準パターン生成処理のフローチャートである。

【0042】標準パターンの生成は、多数話者(Nmax人とする)の発声音声サンプルから各半音節単位の隠れマルコフモデル(HMM; Hidden Markov Model)の確率パラメータを推定することにな

る。

【0043】まず、全ての調音結合パターンを含んだバランス音素テキストを用意し、統計的に十分な多数の話者=Nmax人に発声させて(S38)、AD変換し(S39)、PCM信号を一旦認識ワークRAM3へ格納しておく。次にPCMデータを波形表示あるいはDA変換することで目視あるいは試聴等の作業により、Hmax種類の単音節単位に区切り(S40)、さらに前後の調音結合を考慮して分類したImax種類の半音節単位毎に半音節PCM信号を得る(S41)。この処理の様子を図11の例に示す。

【0044】ここで“K-”の“-”は後方に音が続いていることを示しており、“-A”の“-”は、前方の音に続くことを示している。

【0045】以上までにおいて、半音節種類数×Nmax人分のPCM信号サンプルが準備できたことになる。これらのPCM信号サンプルについてさらに1フレームあたり12ms～16ms程度に分割(フレミング)した(S42)後に、後述の音声分析特徴抽出処理により半音節種類×Nmax人分の特徴ベクトル(フレーム単位)を得る(S43)。

【0046】そして、最終的に“K-”の為の標準パターンを生成するという事は、1つの半音節カテゴリー“K-”に1つのHMMの標準パターンモデルを対応させ、その半音節のFmaxフレーム分の特徴ベクトル出力が対応するHMMの4状態の遷移過程で最も高い確率で出力されるように各状態の確率パラメータおよび状態遷移確率を求めることにある。

【0047】次に、各半音節種類毎に状態数=Jmax個のLeft to Right型HMMの状態出力確率関数Bjの母数とJmax×2個分の状態遷移確率=α(j-1, j)およびα(j, j)の各パラメータを求める方法について説明する。例として、1つの半音節“K-”に対して、図11のような状態数が4の半音節HMMの各パラメータをNmax人の特徴ベクトルから求める場合を説明する。

【0048】ここで、α00～α33は、以下のように状態遷移確率α[*、*]を示す。

・α(j-1, j) : 状態j-1からjへの状態遷移確率(α01、α02、α03)。
・α(j, j) : 状態jからjへの状態遷移確率(α00、α11、α22、α33)。

【0049】また、出力確率=B0～B3は、以下のような算出式になる。

・Bj : 下式の状態jの特徴ベクトル出力確率関数(初期設定する際に多次元正規分布(ガウス)確率密度関数を用いると共に、母数として平均ベクトルおよび共分散行列の各成分を初期値設定する。)

【0050】

【数1】

$$B_j = \sum_{m=0}^{M_{\max}-1} \mu_{mj} \sum_{k=0}^{K_{\max}-1} \lambda_{kmj} \cdot (2\pi |V_{kmj}|)^{-1/2} \cdot \exp[(X_m - \bar{X}_{kmj})^t \cdot V_{kmj}^{-1} \cdot (X_m - \bar{X}_{kmj})]$$

(備考: t は転置操作 (縦ベクトル→横ベクトル)、 V_{kmj} の -1 は逆行列を示す。)

ただし、

j : 状態番号、 $j = 0 \sim J_{\max}-1$

k : 混合分布番号、 $k = 0 \sim K_{\max}-1$ 、 K_{\max} : 混合分布数

m : 特徴ベクトル種類番号、 $m = 0 \sim M_{\max}-1$ 、 M_{\max} : 特徴ベクトル種類数

λ_{kmj} : 混合分布の重みを決める混合分布係数

μ_{mj} : 特徴ベクトル種類間の重みを決める特徴ベクトル重み係数

X_m : 入力音声サンプルのフレーム単位の特徴ベクトル

$$B_j(X) = (2\pi |V_j|)^{-1/2} \cdot \exp[(X - \bar{X}_j)^t \cdot V_j^{-1} \cdot (X - \bar{X}_j)]$$

ここで求めるパラメータは、

【0053】

【数4】

平均ベクトル: \bar{X}_j

共分散行列: V_j と、状態遷移確率: $\alpha(j-1, j)$ および $\alpha(j, j)$ であり、これが半音節 "K-" の標準パターンとなる。

【0054】これらのパラメータは、 N_{\max} 人分の半音節 "K-" の特徴ベクトルサンプルから以下に述べる FB (Forward Backward) アルゴリズム (Baum-welch アルゴリズムともいい、EM (Expectation Maximization) 手法を基本としたアルゴリズム) により反復的に収束するまで演算を繰り返すことにより得られる。

【0055】FB アルゴリズムを述べる前に、まず、半音節 "K-" の N_{\max} 人分の特徴ベクトルを以下のように再定義する。

【0056】○再定義

話者 n の特徴ベクトル: $X \rightarrow X(n, f)$

ただし、

n : 話者番号、 $n = 0 \sim N_{\max}-1$

f : フレーム番号、 $f = 0 \sim F_{\max}(n)-1$

$F_{\max}(n)$: 話者番号 n の半音節 "K-" のフレーム数 (注: 一般に話者毎にサンプルしたフレーム数は異なる)

さらに、以下の FB アルゴリズム処理を行う (S46)。

【0051】

【数2】

\bar{X}_{kmj} : 平均ベクトル

V_{kmj} : 共分散行列

$|V_{kmj}|$: V_{kmj} のノルム (行列式)

なお以降では、説明を容易にするため混合分布数を 1

($K_{\max} = 1$ 、 $\lambda_{kmj} = 1$)、および特徴ベクトル種類を

1 ($M_{\max} = 1$ 、 $\mu_{mj} = 1$) として下式を用いる。

【混合分布数=特徴ベクトル数=1とした場合の出力確率密度関数】

【0052】

【数3】

【0057】 [FB アルゴリズム]

① 共分散行列: V_j 、

【0058】

20 【数5】

平均ベクトル: \bar{X}_j 、

状態遷移確率: $\alpha[j-1, j]$ および $\alpha[j, j]$ の初期値を設定する (S45)。

【0059】 [初期設定値]

【0060】

【数6】

平均ベクトル: $\bar{X}_j \rightarrow \bar{X}_{j_0}$

共分散行列: $V_j \rightarrow V_{j_0}$

状態遷移確率: $\alpha[j-1, j] \rightarrow \alpha[j-1, j]_{_0}$ および $\alpha[j, j] \rightarrow \alpha[j, j]_{_0}$

30 【0061】 ② 半音節 "K-" の HMM に対する前向きパスアルゴリズムによる確率値の目標値 (= FWD_th) と反復処理の最大回数 (= CNT_{max}) を設定する。

【0062】 ③ ④~⑦の処理を $cnt = 1 \sim CNT_{max}$ まで繰り返す。

【0063】 ④ ⑤~⑦の処理を $j = 0 \sim J_{\max}$ ($J_{\max} = 3$) まで繰り返す。

40 【0064】 ⑤ 下式により、各パラメータの更新値を算出する。

【0065】

【数7】

$$\alpha[j-1, j]_{\text{cnt}} = \frac{\sum_{n=0}^{N_{\text{max}}-1} \frac{1}{N_{\text{max}}} \sum_{f=0}^{F_{\text{max}}-1} \text{FWD}(j-1, f-1) \cdot \alpha[j-1, j]_{\text{cnt-1}} \cdot B_{j-1}\{X(n, f)\} \cdot \text{BCK}(j, f)}{\sum_{f=0}^{F_{\text{max}}-1} \text{FWD}(j, f) \cdot \text{BCK}(j, f)}$$

$$\alpha[j, j]_{\text{cnt}} = \frac{\sum_{n=0}^{N_{\text{max}}-1} \frac{1}{N_{\text{max}}} \sum_{f=0}^{F_{\text{max}}-1} \text{FWD}(j, f-1) \cdot \alpha[j, j]_{\text{cnt-1}} \cdot B_j\{X(n, f)\} \cdot \text{BCK}(j, f)}{\sum_{f=0}^{F_{\text{max}}-1} \text{FWD}(j, f) \cdot \text{BCK}(j, f)}$$

$$\bar{X}_{j_{\text{cnt}}} = \frac{\sum_{n=0}^{N_{\text{max}}-1} \frac{1}{N_{\text{max}}} \sum_{f=0}^{F_{\text{max}}-1} \text{FWD}_{\text{s}} \cdot B_j\{X(n, f)\} \cdot \text{BCK}(j, f) \cdot X(n, f)}{\sum_{f=0}^{F_{\text{max}}-1} \text{FWD}(j, f) \cdot \text{BCK}(j, f)}$$

ただし、

$$\text{FWD}_{\text{s}} = \text{FWD}(j-1, f-1) \cdot \alpha[j-1, j]_{\text{cnt-1}} + \text{FWD}(j, f-1) \cdot \alpha[j, j]_{\text{cnt-1}}$$

$$V_{j_{\text{cnt}}} = \frac{\sum_{n=0}^{N_{\text{max}}-1} \frac{1}{N_{\text{max}}} \sum_{f=0}^{F_{\text{max}}-1} \text{FWD}_{\text{s}} \cdot B_j\{X(n, f)\} \cdot \text{BCK}(j, f) - V_{j_{\text{cnt-1}}}}{\sum_{f=0}^{F_{\text{max}}-1} \text{FWD}(j, f) \cdot \text{BCK}(j, f)}$$

ただし、

$$\text{FWD}_{\text{s}} = \text{FWD}(j-1, f-1) \cdot \alpha[j-1, j]_{\text{cnt-1}} + \text{FWD}(j, f-1) \cdot \alpha[j, j]_{\text{cnt-1}}$$

$$V_{j_{\text{cnt-1}}} = \{X_m(n, f) - \bar{X}_{m_{j_{\text{cnt-1}}}}\} \cdot \{X_m(n, f) - \bar{X}_{m_{j_{\text{cnt-1}}}}\}^t$$

【0066】⑥新しいパラメータにより、入力特徴ベクトル $X_m(n, f)$ に対する、HMMモデルの前向きパスアルゴリズムによる出力確率を下式により求める。

【0067】

【数8】

$$\text{出力確率} = \frac{1}{N_{\text{max}}} \sum_{n=0}^{N_{\text{max}}-1} \text{FWD}(J_{\text{max}}, F_{\text{max}})$$

【0068】⑦出力確率 $\geq \text{FWD}_{\text{th}}$ が成立するか（S47）、あるいは、 $\text{cnt} > \text{CNT}_{\text{max}}$ となれば（S49）処理を終了する。

【0069】⑧この時の、パラメータを半音節“K”の標準パターンとする。

【0070】ここで、 $\text{FWD}(j, f)$ は、前向きパスアルゴリズムで求められる確率（Baum-Welchスコアとも呼ばれる）であり、また $\text{BCK}(j, f)$ は、後向きパスアルゴリズムにより求められる確率である。

【0071】また実際には、標準パターンのパラメータを多数サンプルにより求める際は、状態出力確率関数を2つ以上の多次元正規分布の混合分布としたり（例えば男性と女性別等）、特徴ベクトルの種類を増加させる（例えばメルケプストラムベクトルに加えて、フレーム間差分の Δ メルケプストラム、1フレームの平均パワーのフレーム間差分： Δ 平均パワー等）ことでより認識精度を向上可能である。

【0072】以上の処理について、 I_{max} 種類の半音節HMMを標準パターンとして求めて、標準パターンRO

Mに格納しておく（S50）。

【0073】（2）音声分析特徴抽出処理

図12及び図13は不特定話者半音節音声認識器の機能ブロック構成図である。

【0074】図14～図16は不特定話者半音節音声認識処理のフローチャートである。図17はフローチャートの凡例を示す図表である。

【0075】音声分析特徴抽出器の音声分析特徴抽出処理は、以下の全工程をフレーム単位で行う処理である。

【0076】①入力音声のPCM信号を12ms～16ms程度にフレーム分割したフレームPCM信号をブリエンファシス処理（一次差分処理）することにより高周波数帯域を強調する。

【0077】②ブリエンファシス処理済みフレームPCM信号に対し窓処理（ハニング窓等）することによりこの後のFFT処理のためのフレーム境界のスムージング処理をする。

【0078】③窓処理後のフレームPCM信号をN次FFT変換処理することにより線形周波数軸上のN次複素係数ベクトルへ変換する。

【0079】④N次複素係数ベクトルから複素数の絶対値計算によりN次振幅係数ベクトルへ変換する。

【0080】⑤N次振幅係数ベクトルに対し対数演算を施してN次対数振幅係数ベクトルを算出する。

【0081】⑥N次対数振幅係数ベクトルに対し時間軸上への逆離散余弦変換によりP次ケプストラムベクトルを算出する。

【0082】⑦P次ケプストラムベクトルの高時間成分を抑圧する処理（リフタリング）により、声道特性とピッチ特性（声帯特性）を分離し声道特性のみを抽出したP次ケプストラムベクトルを算出する。

【0083】⑧P次声道特性ケプストラムベクトルに対し周波数軸上への離散余弦変換を行い線形周波数軸上におけるN次声道特性対数振幅係数ベクトルに変換する。

【0084】⑨N次声道特性対数振幅係数ベクトルに対しメル周波数軸上（近似的に対数スケール）における等分割点上のスペクトル成分へ補間あるいはスムージング処理を施したベクトル成分を算出することにより人間の聴覚周波数分解能特性（低周波：高→高周波：低）に合わせたN次声道特性対数振幅メル尺度係数ベクトルへ変換する。

【0085】次に、N次声道特性対数振幅メル尺度係数

$$D_{i,j} = \sum_{m=0}^{Nm-1} \sum_{k=0}^{Nk-1} \lambda_{i,j,k,m} \cdot (|V_{i,j,k,m}| + (\Delta X_{i,j,k,m}) \cdot (V_{i,j,k,m}) \cdot \Delta X_{i,j,k,m})$$

（備考：tは転置操作（縦ベクトル→横ベクトル）、V_{i,j,k,m}の-1は逆行列を示す。）

X_{in}：入力音声の特徴ベクトル

X_{i,j,k,m}：標準パターンの特徴平均ベクトル

D_{i,j}：半音節 = i、状態 = j の状態の状態距離計算値

V_{i,j,k,m}：標準パターンの特徴量共分散行列

|V_{i,j,k,m}|：共分散行列V_{i,j,k,m}のノルム（分散値）

i：半音節番号、i = 0 ~ I_{max} - 1、I_{max}：全半音節種類

j：1半音節のHMMにおける状態番号、j = 0 ~ J_{max} - 1、J_{max}：1HMMの全状態数

k：混合分布番号、k = 0 ~ K_{max} - 1、K_{max}：混合分布数

m：特徴ベクトル種類番号、m = 0 ~ M_{max} - 1、M_{max}：総特徴ベクトル種類数

【0090】（4）Viterbi処理（パターンマッチング処理）

例として、以下のような氏名リストの認識を行うことを考える。“たかはし”という単語はまず、“TAKAHASI”と母音、子音列に変換され、更に“T-, -A-, -K-, -A-, -H-, -A-, -S-, -I-”という半音節列に規則的に、分解される。

（半音節列）“たかはし”

（母音、子音列）“TAKAHASI”

（半音節列）

“T-, -A-, -K-, -A-, -H-, -A-, -S-, -I-”

各半音節は、前述の図11のように標準パターンHMMを持って表現されていた。

【0091】これにより、“たかはし”という単語のHMM連結モデルは図18のようになる。

【0092】このHMM連結モデルから、一種の累積確

ベクトルに対し時間軸上への逆離散余弦変換によりQ次メルケプストラムベクトルへ変換する。

【0086】以上により、1フレーム分PCM信号から入力特徴ベクトルが得られる（S58）。

【0087】（3）状態距離計算

状態距離計算器は、入力音声のアナログ信号をAD変換し、さらに1/2ms～16ms程度にフレーム分割したPCM信号から音声分析特徴抽出器の音声分析特徴抽出処理により得られた入力特徴ベクトルと、標準半音節HMMの全状態の標準特徴ベクトルとの状態距離値を下記の距離計算式により算出する（S60）。

【0088】[距離計算式]

$$\Delta X_{i,j,k,m} = X_{in} - X_{i,j,k,m}$$

【0089】

[数9]

率を算出するのがViterbiアルゴリズムである。

【0093】Viterbiアルゴリズムは、基本的に図19の最適パス選択処理の繰り返しである。

【0094】まず分かりやすい例としてとして、1つの半音節HMM“T-”のViterbiスコア算出例を図20に示す。又、入力パターンはF_{max}フレーム分とする。図20の例のように、各フレーム入力毎に、全状態のViterbiスコアを求めていき、全フレーム分について、算出した時の状態3のViterbiスコアが、“T-”の入力特徴ベクトルに対する標準パターンの出力確率となる。さて、これを“たかはし”という単語のHMM連結モデルに適用する場合には、状態数が、4×8=32、入力フレーム数=F_{max}として、半音節“-I”のHMMにおける状態3のViterbiスコアを算出することで、入力特徴ベクトル“たかはし”のHMM連結モデルからの出力確率が求まることになる。

【0095】実際には、認識対象リストが、“たかはし”の他にも複数存在するので、例えば“いとう”という氏名に対しても同様の半音節列への変換をしてViterbiスコアを算出する（S75）。そして、全認識対象リストにおけるViterbiスコアから最も確率値の高い（距離値の小さい）認識対象リストの1つを認識結果とする（S86）。以下にViterbiアルゴリズムの処理手順を示す。

【0096】[Viterbiアルゴリズム]

①（S57～S81）②～⑥の処理をf=0～F_{max}-1まで繰り返す。

②（S73～S79）③～⑥の処理をw=0～W_{max}-1まで繰り返す。

③（S73～S77）④～⑥の処理をs=0～S_{max}(w)-1まで繰り返す。

④（S73）i←[状態=sが属する半音節番号]

$j \leftarrow$ [状態 = s が属している半音節番号 = i における HMM 内の状態番号]

⑤ (S74) $Path(j-1, j) = \alpha(j-1, j) + G(w, s-1)$

$Path(j, j) = \alpha(j, j) + G(w, s)$

⑥ (S75) 累積距離値: $G(w, s) = \text{Max}[Path(j-1, j), Path(j, j)] + Dij$ の計算

ただし、

f : 入力フレーム番号、 F_{max} : 全フレーム数

w : 認識対象 (氏名 or 電話番号) リストの番号、 W_{max} : 全リスト数

s : リスト内状態連結の通し番号、 $S_{\text{max}}(w)$: 認識対象リスト = w 番の全連結状態数

$\alpha(j-1, j)$: i 番の半音節 HMM において状態 $j-1$ から j への状態遷移距離値 ($j-1 < 0$ の場合は距離値 = 0)

$\alpha(j, j)$: i 番の半音節 HMM において状態 j から j への状態遷移距離値

Dij : 半音節 = i 、状態 = j の状態の入力特徴ベクトルとの状態距離値

[0097] なお、Viterbi スコアは確率値として説明していたが、実際には桁数の制限等でアンダーフロー等の問題を避ける為に、対数演算を施したもので Viterbi スコアを計算することもある。

[0098] 又、演算量低減及びメモリ量低減の為に標準パターンの平均ベクトル、共分散行列をクラスタリングして演算量を低減する工夫もある。

[0099] 例えば、半音節 HMM が 250 種類あると、 $250 \times 4 = 1000$ 種類の平均ベクトルと共分散行列を用意しなければならないが、例えば平均ベクトルを 512 カテゴリー (この場合、共分散行列も 512 種類) に、また分散行列のノルムを 256 カテゴリーに、ベクトル量子化の手法 (セントロイドベクトル等) によりベクトル値を代表させることで、クラスタリングを行うと、演算量とメモリ量が $1/2 \sim 1/4$ になる。実際にこのような工夫により、認識性能は劣化することなく演算量とメモリ量を低減することは可能である。

[0100] 4. 音声登録用検索器

音声登録用検索器 (以降検索器という) は、不特定話者半音節認識器 (以降認識器という) から 1 音節単位の認識結果を第 1 候補 $\sim N$ 候補まで距離値と共に情報受信し、データベース ROM1 に格納されている日本人全氏名の平仮名文字データリスト内で、1 音節単位に発声された氏名の音声信号から最も距離の近い氏名の文字コードを検索し、これを認識音節列ワーク RAM へ格納する。これにより、音声信号入力による氏名登録が行われることになる。以降では、この音声登録動作について具体的な実施例を説明する。

[0101] まず、発声される音声信号を “た” +

“か” + “は” + “し” と、1 音節単位に離散発声されたものを例とする。認識器から検索器には、1 音節の発声毎に認識結果の文字コードと距離値が図 20 の例のように複数候補出力される。

[0102] ここで、第 1 候補が実際の発声音声 =

“た” に対し “な” になっているのは、1 音節分の発声方法が似かよっているためである。これは、基本的に子音 + 母音という 1 音節の構造上子音だけが異なり、母音が全て同じ場合には、たとえ人の聴覚識別能力であっても間違って聞き取ってしまうものと同じ事である。認識器では、メルケプストラムという音声信号の特徴量を抽出しているがこれは声道の特徴量を抽出していることと同じ意味である。

[0103] “た” の PCM 信号を実際に認識器により認識処理した場合した場合は、“T” + “-A” という半音節 HMM 結合より出現される確率が最も高い (距離値が小さい) のではなく、実際には人間の微妙な発声変形等の影響により、“子音の半音節” + “-A” もほぼ同等の認識距離となってしまうのである。例えば同母音系の “な” を考えてみると、半音節は “N” + “-A” となり、母音部は同じとなるため、また子音部についても “た” と “な” については、どちらも子音を発声する瞬間は舌を上あごにつけてから “-A” を発声するため、“た” と誤認識し易くなってしまうのである。しかし、複数音節から通常構成される氏名の場合は、上例において残る “か”、“は”、“し” の音節の認識結果も誤認識し易くなるのは変わらないが、累積確率を存在する氏名の音節列のみに対して計算することにより、単音節認識の誤認識を補うことが出来るのである。これを以降に示す。

[0104] まず、“た”、“か”、“は”、“し” の各発声に対する単音節認識結果が図 22 の例のようになったとする。

[0105] まず、検索装置ではこの情報を登録用音節列ワーク RAM へ一時格納しておく。次に、発声回数が 4 回であることがカウンタ等によりカウント出来るため、データベース ROM から四文字の氏名リストを検索し、それらもワーク RAM に格納しておく。次に、4 文字氏名リストの文字コードを調べて認識結果の全ての文字コードが 1 つでも含まれている氏名を絞り込む。更に、絞り込んだ 4 文字氏名リストについて認識結果の距離値でもって累積距離値を累加演算処理により計算していく。ここで、ある氏名のある文字コードには距離がないものがあるが、以下の方法により計算する。

[0106] ここで、また例として絞り込まれた 4 文字氏名リストが図 23 のようになったとする。

[0107] 次に、これらの候補氏名に対して累積距離値を算出する。この際認識結果リストにある候補に対しては、発声順も考慮して距離値を加算していき、認識結果リストにない文字に対しては最大距離値 = 5.0 を設

定して累加算演算を行う。図 23 の例の一覧により、“た” + “か” + “は” + “し” が最も小さい値になり、これが登録氏名となる。

【0108】

【発明の効果】以上説明したように本発明は、入力音声に対する、登録済み氏名又は電話番号の単音節コードデータ列と半音節単位の標準パターンデータとを用いた認識処理により、携帯電話機等における氏名と電話番号登録操作、およびダイヤル操作等のキー操作における煩雑さを低減することができ、音声認識ダイヤル用の氏名と電話番号の登録操作時間が従来例と比べて短縮され、また操作間違いも少なくなるという効果がある。

【0109】また、従来技術の認識自動ダイヤル電話機では、明記されていなかった音声から文字表示を行う音声認識処理について、この技術の実現手段を明記することで本発明の属する技術分野において実際に利用可能となるという効果がある。

【図面の簡単な説明】

【図 1】本発明の実施の形態の全体構成を示すブロック図である。

【図 2】音声登録動作モードのフローチャートである。

【図 3】不特定話者半音節音声認識器により得られた 4 音節 × 5 候補分の認識結果を示す図である。

【図 4】電話番号登録動作モードのフローチャートである。

【図 5】電話番号の認識結果の数字を直接表すコードへの変換を示す図である。

【図 6】音声ダイヤル動作モードのフローチャートである。

【図 7】標準パターン生成器の機能ブロック構成図である。

【図 8】標準パターン生成器の機能ブロック構成図である。

【図 9】標準パターン生成処理のフローチャートである。

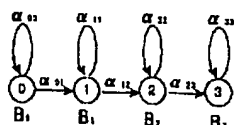
【図 10】単音節 PCM 信号から半音節 PCM 信号を得る例の図である。

【図 11】状態数が 4 の半音節 HMM を示す図である。

【図 12】不特定話者半音節音声認識器の機能ブロック構成図である。

【図 11】

(例) 状態数 = 4 ($J_{max} = 4$) の HMM



【図 13】不特定話者半音節音声認識器の機能ブロック構成図である。

【図 14】不特定話者半音節音声認識処理のフローチャートである。

【図 15】不特定話者半音節音声認識処理のフローチャートである。

【図 16】不特定話者半音節音声認識処理のフローチャートである。

【図 17】フローチャートの凡例を示す図表である。

【図 18】“たかはし”という単語の HMM 連結モデルを示す図である。

【図 19】最適パス選択処理を示す図である。

【図 20】1つの半音節 HMM “T” の Viterbi スコア算出例を示す図である。

【図 21】不特定話者半音節音声認識器から音声登録用検索器に複数候補出力される、1 音節の発声毎の認識結果の文字コードと距離値を示す図である。

【図 22】“た、”、“か”、“は”、“し”の各発声に対する単音節認識結果を示す図である。

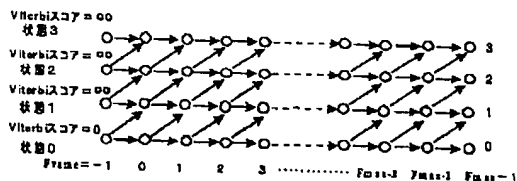
【図 23】絞り込まれた 4 文字氏名リストに対して累積距離値を算出する図である。

【符号の説明】

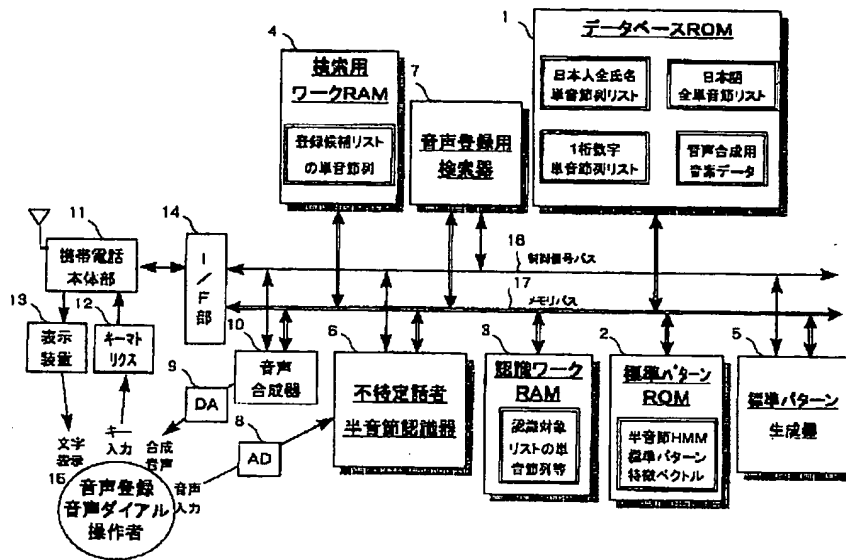
- 1 データベース ROM
- 2 標準パターン ROM
- 3 認識ワーク RAM
- 4 検索ワーク RAM
- 5 標準パターン生成器
- 6 不特定話者半音節音声認識器
- 7 音声登録用検索器
- 8 ADコンバータ
- 9 DAコンバータ
- 10 音声合成器
- 11 携帯電話本体部
- 12 キーマトリクス
- 13 表示装置
- 14 I/F部
- 15 音声登録音声ダイヤル操作者
- 16 制御信号バス
- 17 メモリバス

【図 20】

(例) “T” の Viterbi スコア計算方法



【図1】

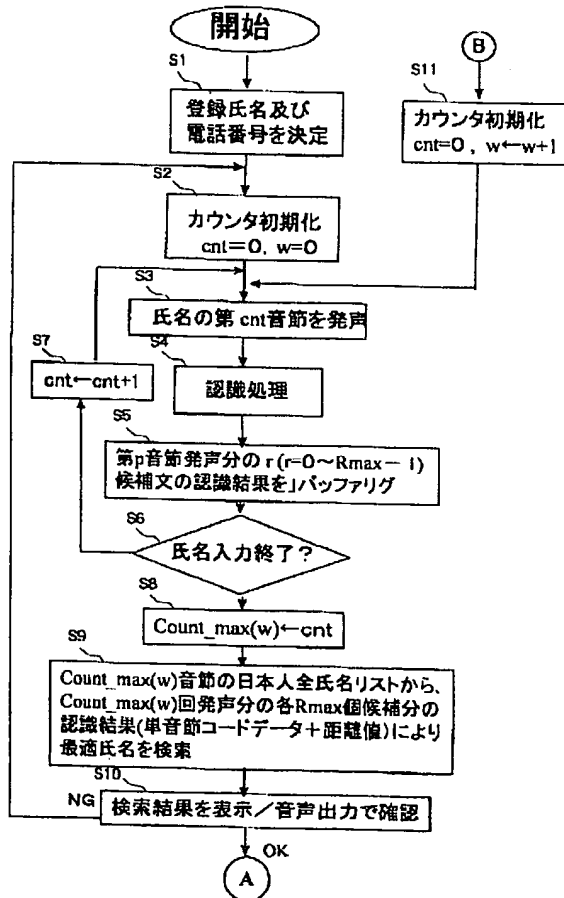


【図5】

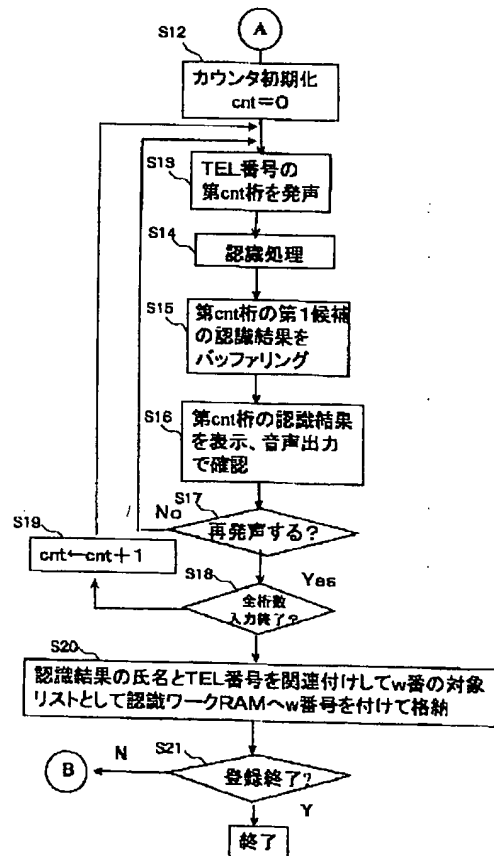
認識結果 数字表記

「せ+ろ」	→	「0」
「い+ち」	→	「1」
「に」	→	「2」
「さ+ん」	→	「3」
「よ+ん」	→	「4」
「ご」	→	「5」
「ろ+く」	→	「6」
「な+な」	→	「7」
「は+ち」	→	「8」
「さ+ゆ+う」	→	「9」

【図2】



【図4】



【図3】

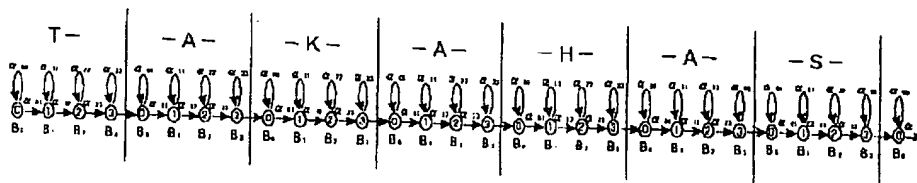
(例)

	第1候補	第2候補	第3候補	第4候補	第5候補
第1発声：“た”	な (1.5)	た (1.7)	ら (2.6)	さ (3.5)	か (4.5)
第2発声：“か”	あ (1.5)	は (1.7)	か (1.8)	や (3.5)	わ (4.5)
第3発声：“は”	あ (1.4)	は (1.5)	か (1.7)	や (3.0)	わ (4.5)
第4発声：“し”	い (1.4)	し (1.5)	ひ (1.6)	ち (2.0)	じ(ぢ) (3.0)

(注) 括弧内の数値は不特定話者半音節音声認識器から出力される距離値。

【図18】

(例) “たかはし”のHMM連結モデル



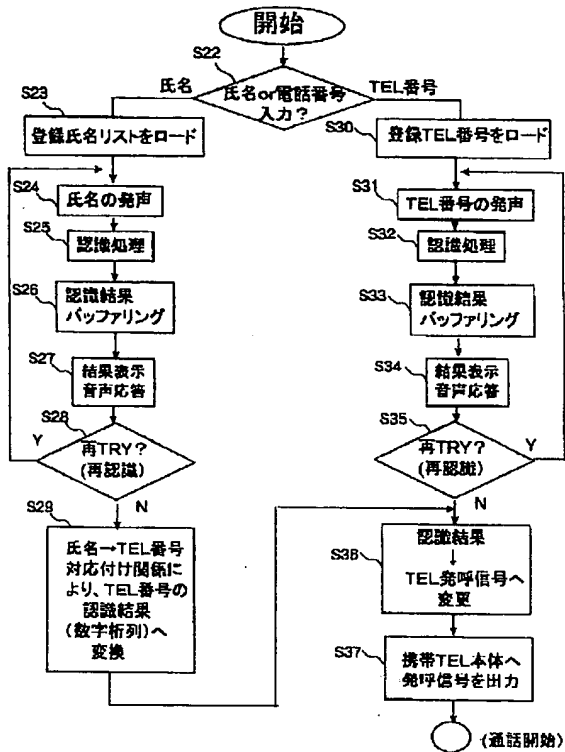
【図21】

(例) 第1発声＝“た”に対する1音節認識結果

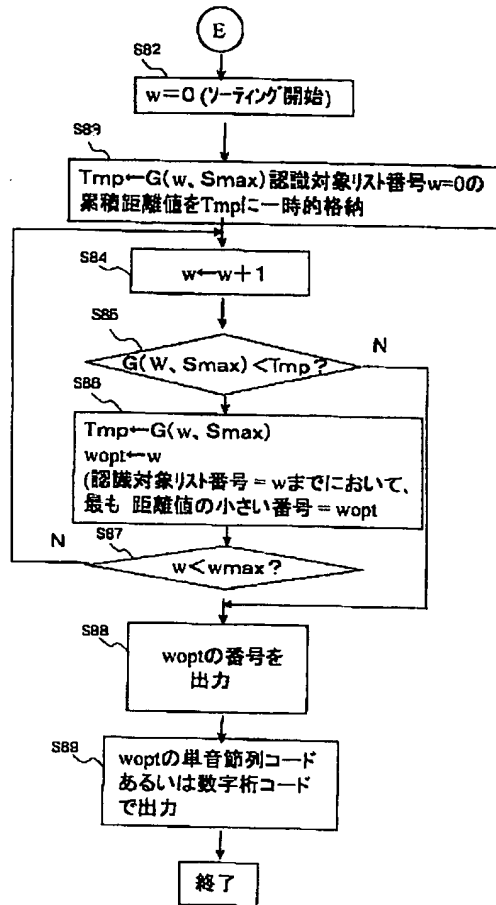
候補	第1位	第2位	第3位	第4位	第5位
音節	な	た	ら	さ	か
距離値	1.5	1.7	2.6	3.5	4.5

(注) 実際は、音節は音節コード値。距離値範囲：1.0～5.0。

【図6】

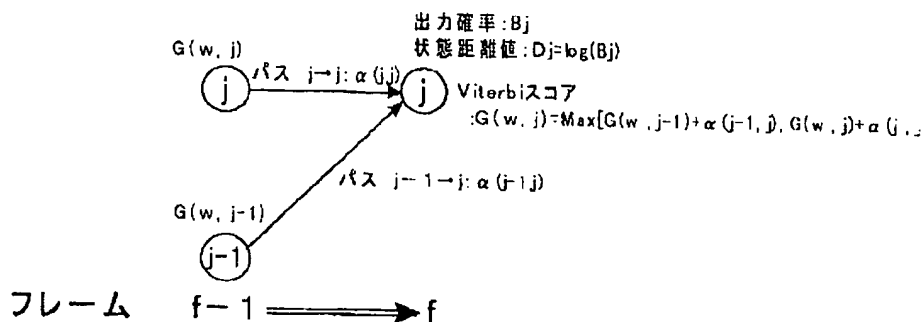


【図16】

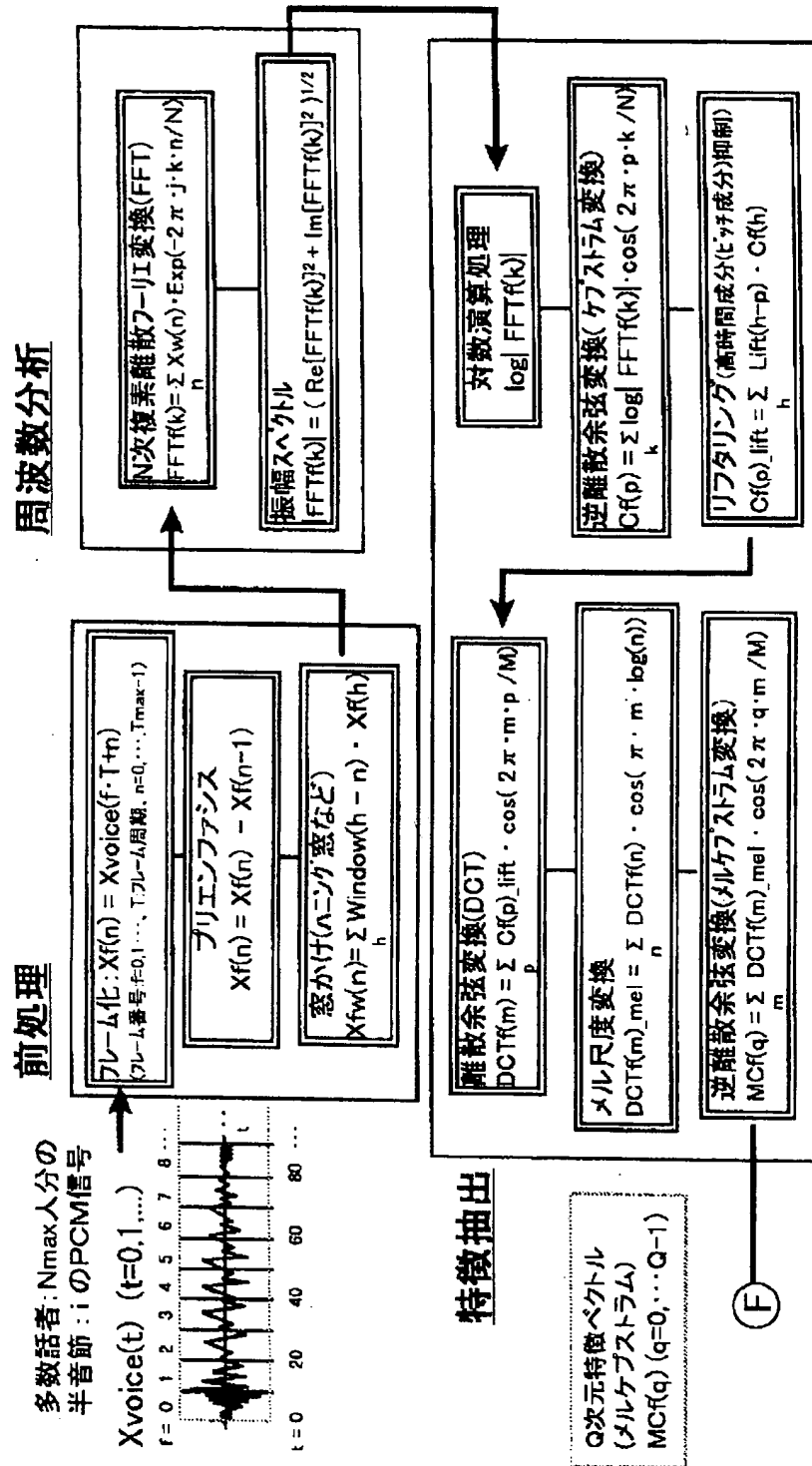


【図19】

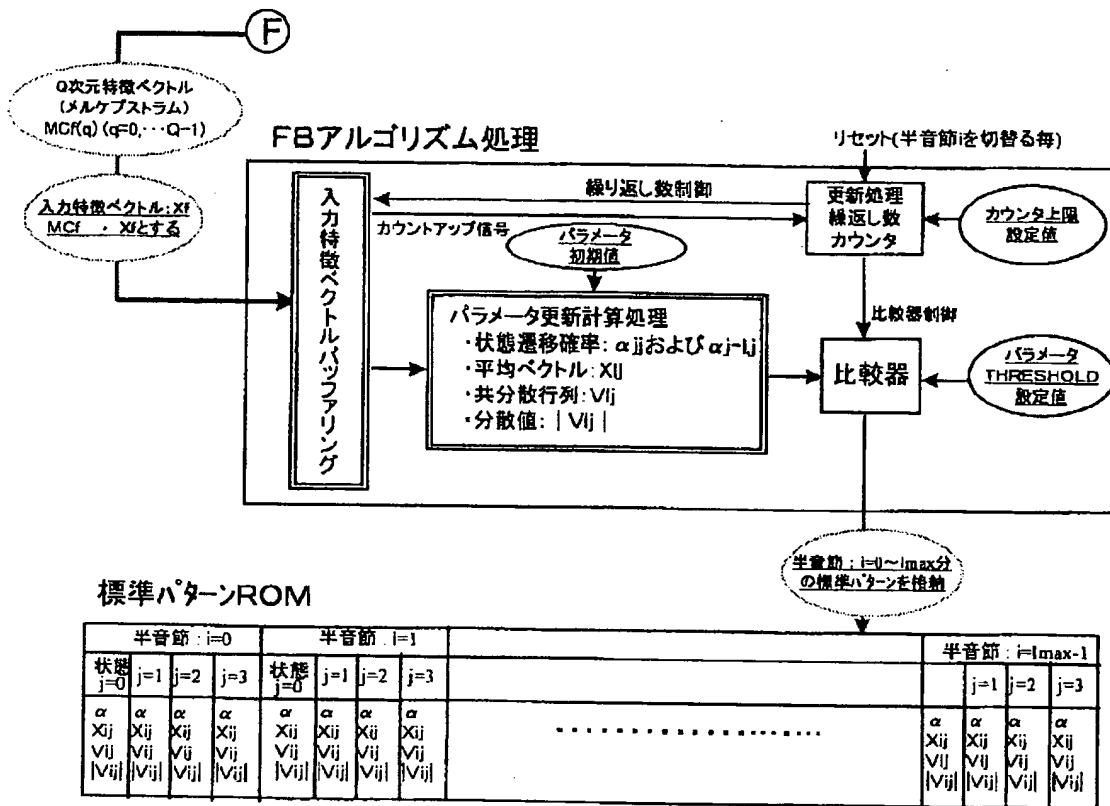
(例) 最適パスの選択方法



【図7】



【図8】



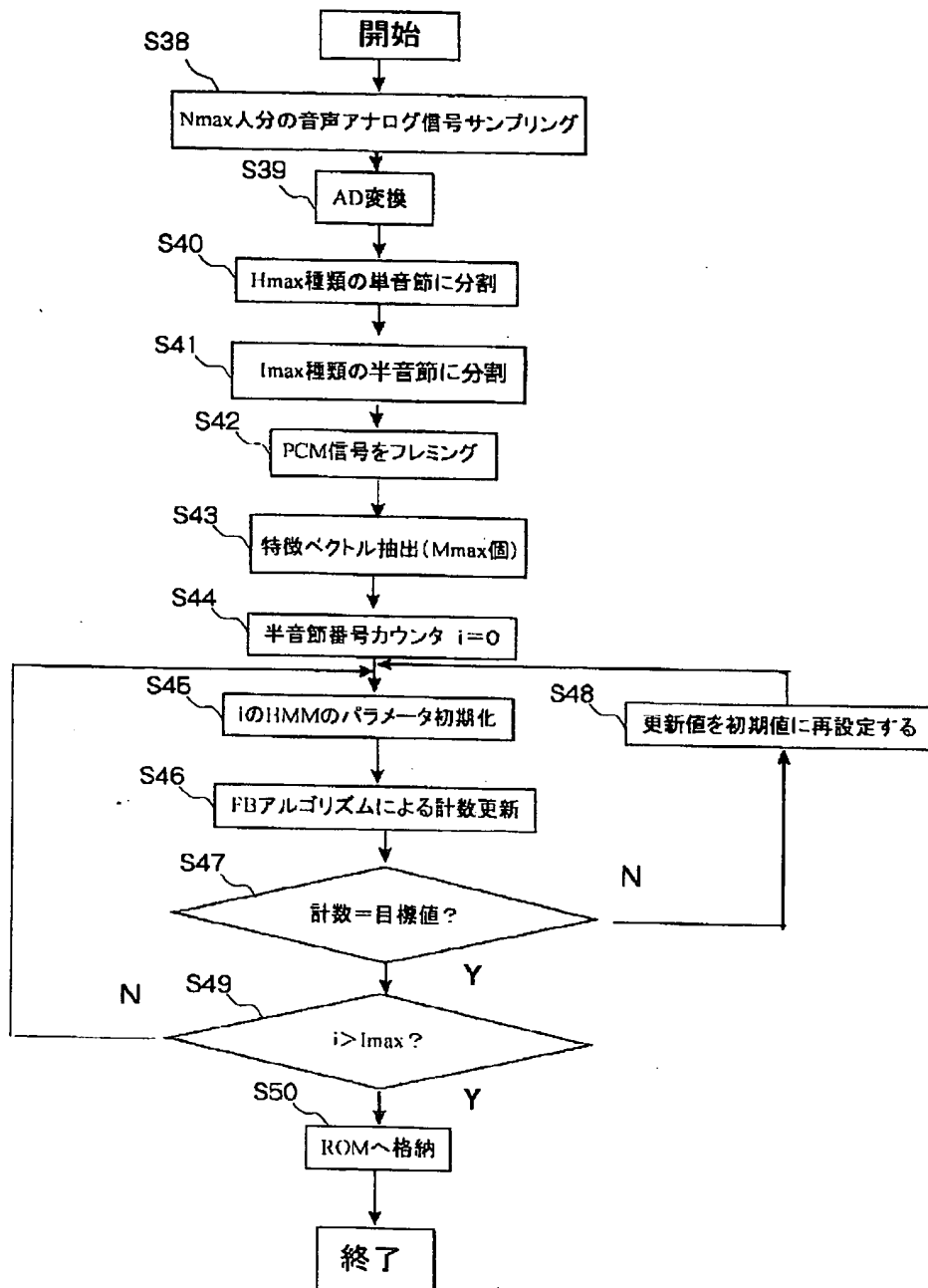
(注) i: フレーム番号, j: 半音節番号, j: 状態番号 平均ベクトル: X_{ij} 、共分散行列: V_{ij} 、分散値: $|V_{ij}|$

【図22】

(例)

	第1候補	2	3	4	5
発声“た”	な (1.5)	た (1.7)	ら (2.6)	さ (3.5)	か (4.5)
“か”	あ (1.5)	は (1.7)	か (1.8)	や (3.5)	わ (4.5)
“は”	あ (1.4)	は (1.5)	か (1.7)	や (3.0)	わ (4.5)
“し”	い (1.4)	し (1.5)	ひ (1.6)	ち (2.0)	じ(ち) (3.0)

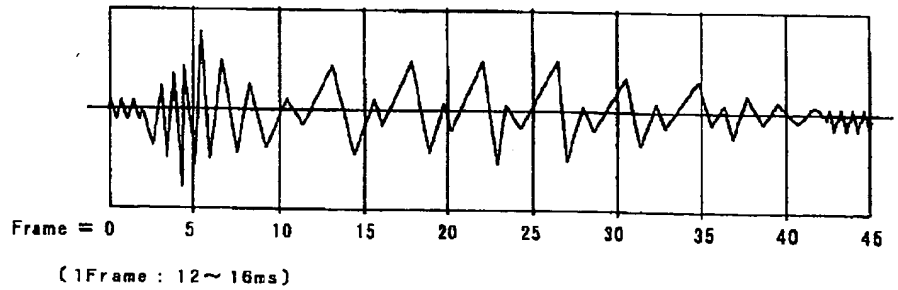
【図9】



【図 10】

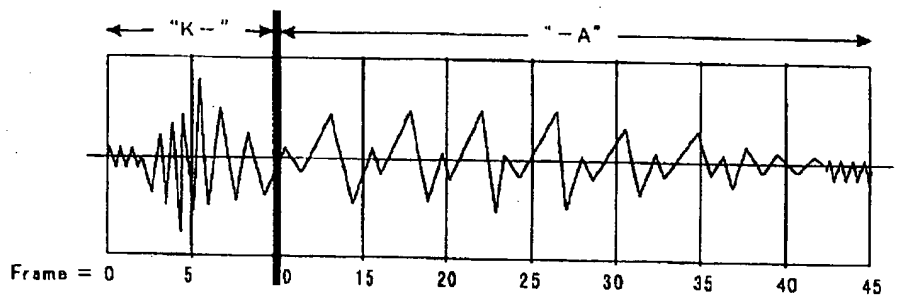
(例)ある話者の発声音：単音節“KA”のPCM信号波形

〔単音節“KA”(か)のPCM信号波形〕

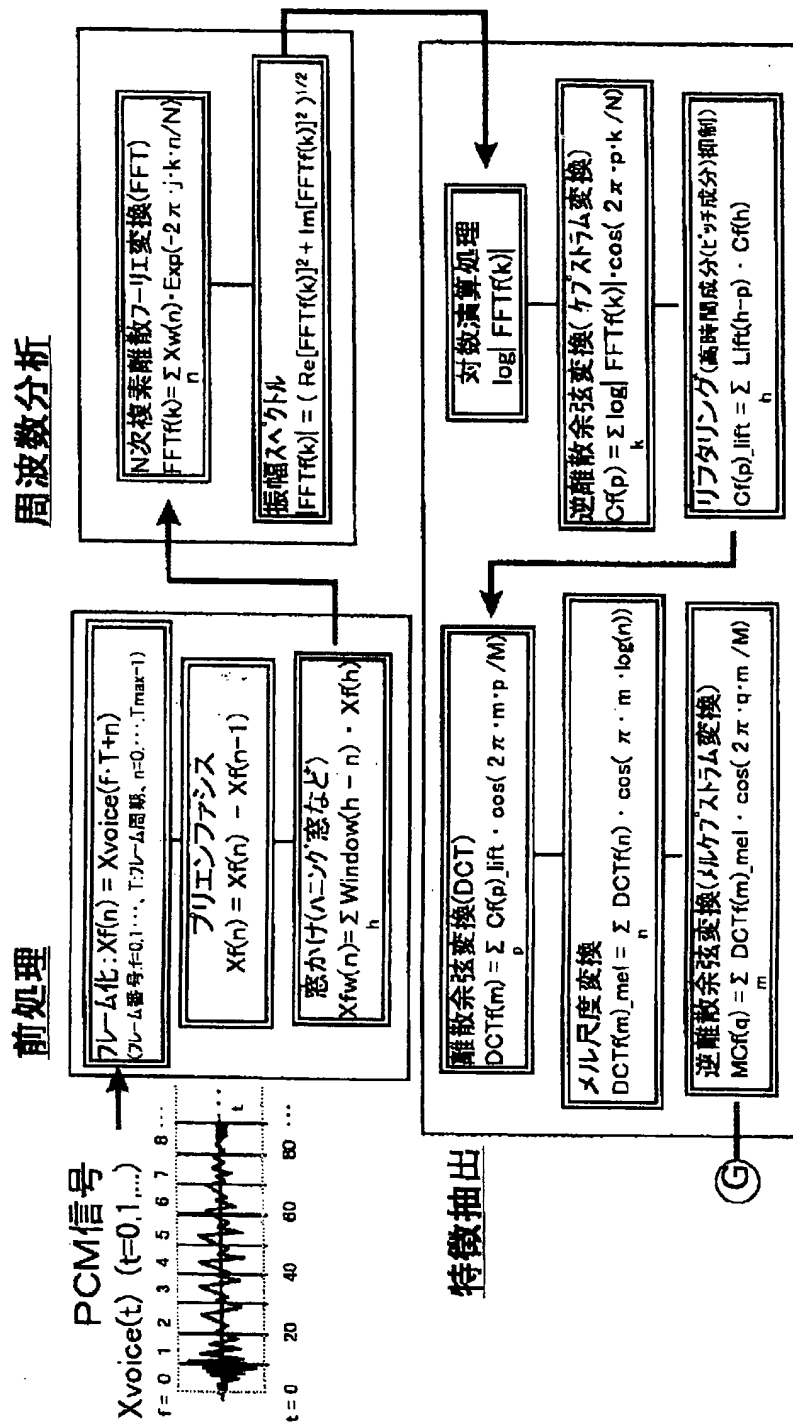


↓

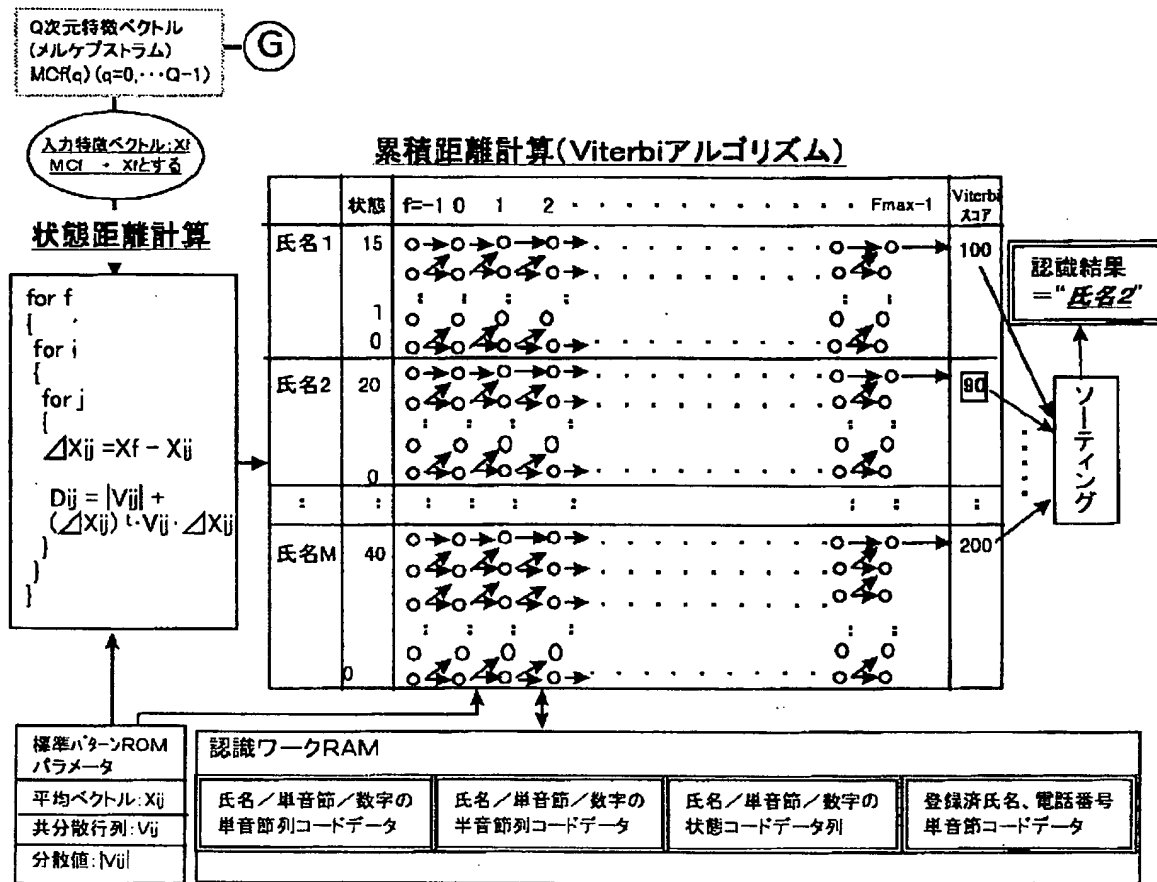
〔半音節“K-”+“-A”に分割されたPCM信号波形〕



【図12】

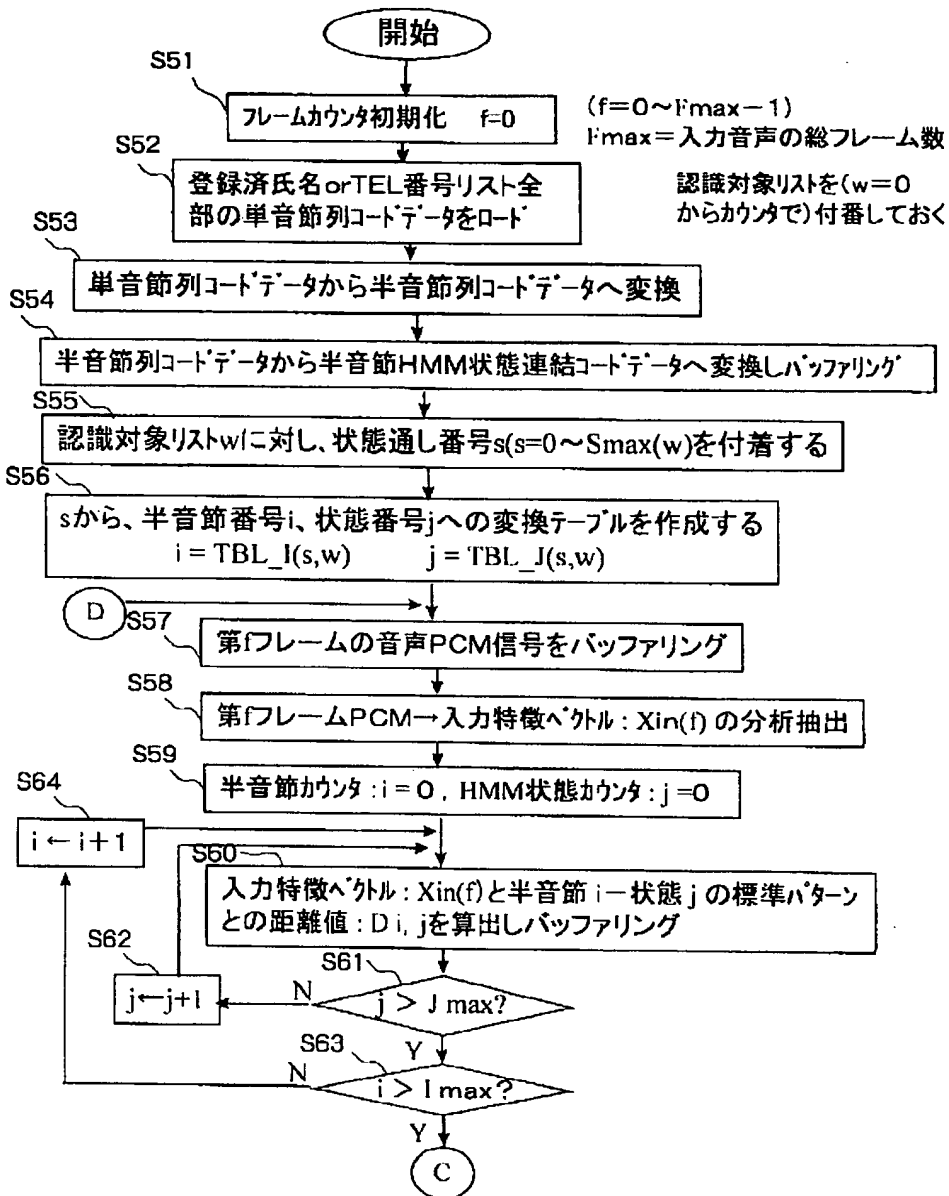


【図13】

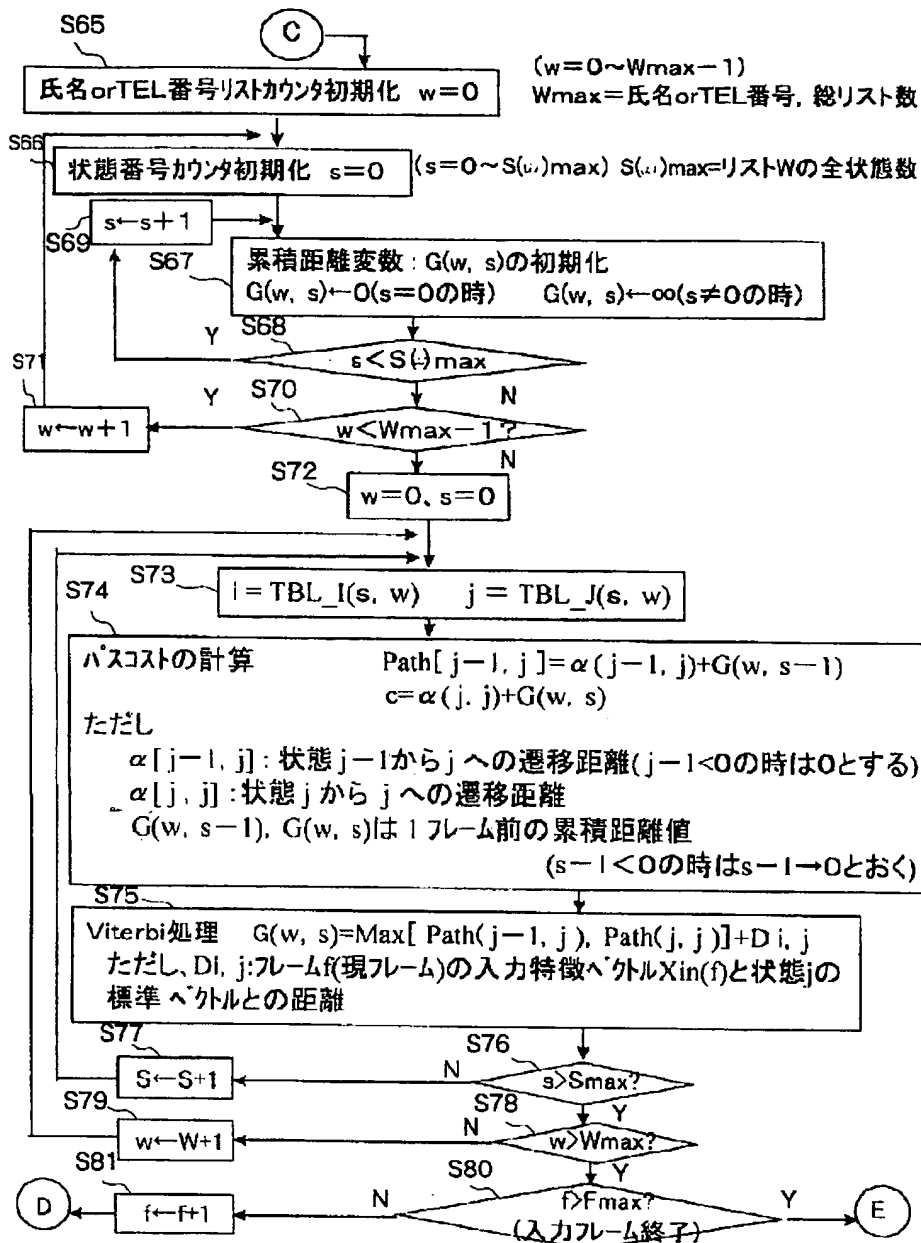


(注) F: フレーム番号, i: 半音節番号, j: 状態番号

【図14】



【図15】



【図17】

	Index	値域
標準話者数	n	$0 \leq n \leq N_{\max} - 1$
単音節番号	h	$0 \leq h \leq H_{\max} - 1$
半音節番号	i	$0 \leq i \leq I_{\max} - 1$
状態番号 (半音節番号 i 内の状態番号)	j	$0 \leq j \leq J_{\max} - 1$
認識対象リスト番号 (氏名あるいはTELリストなど)	w	$0 \leq w \leq W_{\max} - 1$
フレーム番号	f	$0 \leq f \leq F_{\max} - 1$
サンプリング時間	t	$-\infty \leq t \leq \infty$
連結状態番号 (リスト w 内の連結状態の通し番号)	s	$0 \leq s \leq S_{\max}(w) - 1$
混合分布	k	$0 \leq k \leq K_{\max} - 1$
特徴ベクトル種類	m	$0 \leq m \leq M_{\max} - 1$
認識結果候補カウンタ	r	$0 \leq r \leq R_{\max}$
汎用カウンタ	cnt	$0 \leq cnt \leq \infty$

【図23】

(例)	第1発声	2	3	4 累積距離値
1	"あ" (5)	"か" (1.8)	"ざ" (5)	"わ" (5) → 18.8
2	"か" (4.5)	"き" (5)	"さ" (5)	"わ" (5) → 19.5
3	"さ" (3.5)	"わ" (4.5)	"は" (1.5)	"し" (1.5) → 11.0
4	"た" (1.7)	"き" (5)	"ざ" (5)	"わ" (5) → 16.7
5	"た" (1.7)	"か" (1.8)	"は" (1.5)	"し" (1.5) → <u>6.5</u>
6	"あ" (1.5)	"か" (1.8)	"ざ" (5)	"わ" (5) → 13.3
7	"は" (1.5)	"し" (1.8)	"か" (5)	"わ" (5) → 13.3
8	"ま" (1.5)	"え" (1.8)	"か" (5)	"わ" (5) → 13.3
9	"み" (1.5)	"つ" (1.8)	"は" (5)	"し" (5) → 13.3
10	"や" (1.5)	"ま" (1.8)	"か" (5)	"わ" (5) → 13.3